



On the sequential construction of optimum bounded designs

Luc Pronzato

► To cite this version:

Luc Pronzato. On the sequential construction of optimum bounded designs. Journal of Statistical Planning and Inference, 2006, 136, pp.2783 - 2804. 10.1016/j.jspi.2004.10.020 . hal-00416063

HAL Id: hal-00416063

<https://hal.science/hal-00416063>

Submitted on 11 Sep 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the sequential construction of optimum bounded designs

Luc Pronzato

*Laboratoire I3S, CNRS-UNSA, Bât. Euclide, Les Algorithmes,
2000 route des Lucioles, BP 121, 06903 Sophia Antipolis cedex, France*

Abstract

We consider a parameter estimation problem with independent observations where one samples from a finite population of independent and identically distributed experimental conditions X . The size of the population is N but only n samples, a proportion α of N , can be used. The quality of a sample is measured by a regular optimality criterion $\phi(\cdot)$ based on the information matrix, such as the D -criterion. The construction of an optimal approximate design bounded by μ/α , with μ the probability measure of X , can be used to construct a sampling strategy which is asymptotically optimum (when the size N of the population tends to infinity). We show that a sequential strategy which does not require any information on μ is also asymptotically optimum. Some possible applications are indicated.

Key words: sequential design, sampling, constrained design measure, bounded design.

1991 MSC: 62K05, 62L05

Email address: pronzato@i3s.unice.fr (Luc Pronzato).

URL: <http://www.i3s.unice.fr/pronzato/> (Luc Pronzato).

1 Introduction

We consider a parameter estimation problem, with θ the vector of parameters to be estimated and X the experimental variables, $X \in \mathcal{X} \subset \mathbb{R}^q$. We assume that the observations are independent, so that the Fisher information matrix is the sum of rank-one matrices of the type $f(X)f^\top(X)$. For instance, this may correspond to estimation in a linear regression model, with independent observations $Y_k = f^\top(X_k)\bar{\theta} + \varepsilon_k$, where the ε_k 's are independently identically distributed (i.i.d.) with $\mathbb{E}\{\varepsilon_k\} = 0$ and $\bar{\theta} \in \mathbb{R}^d$ is the unknown true value of the model parameters to be estimated.

We shall assume that $f(\cdot)$ is continuous on \mathcal{X} . We consider design criteria $\Phi(\cdot)$ that are functions of the information matrix \mathbf{M} , with $\Phi[\mathbf{M}]$ to be maximized, and generalized designs ξ that are probability distributions on the set \mathcal{X} . We denote Ξ the set of such designs,

$$\mathbf{M}(\xi) = \int_{\mathcal{X}} f(x)f^\top(x)\xi(dx),$$

and $\phi(\xi) = \Phi[\mathbf{M}(\xi)]$, $\xi \in \Xi$. We give in Appendix A a list of assumptions on Φ that will be used throughout the paper. We shall always assume that Φ is strictly concave (**H_Φ1**), differentiable (**H_Φ2**) and increasing (**H_Φ3**). The assumptions are discussed in the same appendix; they are satisfied in particular when $\Phi(\mathbf{M}) = \log \det \mathbf{M}$ or $\Phi(\mathbf{M}) = -\text{trace}(\mathbf{A}\mathbf{M}^{-p})$, with p a positive integer and \mathbf{A} a positive definite matrix.

We consider the situation where the experimental conditions $X_k \in \mathcal{X}$ form a sequence of i.i.d. variables sampled in a population of size N , and only $n < N$ samples can be used. We focus on the sequential problem, where, as soon as a new sample X_k becomes available, we must decide whether to use it or not,

that is, to observe Y_k or not. Notice the difference with a standard experimental design problem where the X_k 's can be *chosen*: here we can only decide to accept or reject X_k . The paper is rather theoretically oriented, but many practical decision problems could be formulated in this way. For instance, for some experiments in nuclear physics events are selected according to the energy dissipated in a detector (C.E.R.N., 1994), but the selection could be based on the information content of the event, measured by its contribution to the information matrix; in phase-I clinical trials, the volunteers could be selected according to their size, weight, age, etc., all variables to be used to build a model for the tolerance dose. Other possible developments are given in Section 6. Notice that the case $\dim(\theta) = 1$ corresponds to a variant of the secretary problem, see (Pronzato, 2001a).

Let μ denote the probability measure of X_1 , with $\int_{\mathcal{X}} \mu(dx) = 1$. We assume that μ is such that

$$\mathbf{M}(\mu) = \mathbf{E}\{f(X_1)f^\top(X_1)\} = \int_{\mathcal{X}} f(x)f^\top(x)\mu(dx)$$

exists, with $-\infty < \phi(\mu) < \infty$; a list of additional assumptions on μ is given in Appendix A. The sequence of decisions will be denoted (u_k) : $u_k = 1$ if we decide to observe Y_k , with experimental conditions X_k , and $u_k = 0$ otherwise, with, for any admissible policy,

$$u_j \in \mathcal{U}_j \subseteq \{0, 1\}, \quad j = 1, \dots, N, \quad \sum_{j=1}^N u_j = n. \quad (1)$$

(Note that X_k is known when u_k is chosen.) The associated information matrix $\mathbf{M}_{N,n}$ (normalized, per sample collected) is given by

$$\mathbf{M}_{N,n} = \frac{1}{n} \sum_{k=1}^N u_k f(X_k)f^\top(X_k). \quad (2)$$

(It may be singular for small n , but $\Phi(\mathbf{M}_{N,n}) > -\infty$ for n large enough if the X_k 's are selected in a suitable population, see $\mathbf{H}_\Phi 1$, $\mathbf{H}_\mu 1$.)

For N finite, we shall consider the following sequential problem

$$\text{maximise } \mathbf{E}\{\Phi(\mathbf{M}_{N,n})\} \quad (3)$$

with respect to (u_j) satisfying (1), the expectation $\mathbf{E}\{\cdot\}$ being with respect to the product measure $\mu^{\otimes N}$ of X_1, \dots, X_N (we shall see, c.f. (12), that the concavity and increasing properties of Φ and $\phi(\mu) < \infty$ imply that $\mathbf{E}\{\Phi(\mathbf{M}_{N,n})\} < \infty$ for any N, n , $0 < n \leq N$, and any sequence (u_j)).

For any sequence (u_j) and any step k , $1 \leq k \leq N$, a_k will denote the number of observations already made; that is,

$$a_k = \sum_{j=1}^{k-1} u_j, \quad (4)$$

with $a_1 = 0$. For each $k \in \{1, \dots, N\}$, the optimal decision at step k is obtained by solving:

$$\begin{aligned} & \max_{u_k \in \mathcal{U}_k} [\mathbf{E}_{X_{k+1}} \{ \max_{u_{k+1} \in \mathcal{U}_{k+1}} [\mathbf{E}_{X_{k+2}} \{ \max_{u_{k+2} \in \mathcal{U}_{k+1}} [\dots \\ & \mathbf{E}_{X_{N-1}} \{ \max_{u_{N-1} \in \mathcal{U}_{N-1}} [\mathbf{E}_{X_N} \{ \max_{u_N \in \mathcal{U}_N} [\Phi([1/n] \sum_{i=1}^N u_i f(X_i) f^\top(X_i))] \} \dots] \} \} \} \} \} \}, \quad (5) \end{aligned}$$

where $\mathbf{E}_{X_j}\{\cdot\}$ denotes the expectation with respect to X_j , distributed with the measure μ , and the sets \mathcal{U}_j satisfy

$$\mathcal{U}_j = \mathcal{U}_j(a_j) = \begin{cases} \{0\} & \text{if } a_j = n, \\ \{1\} & \text{if } a_j + N - j + 1 = n, \\ \{0, 1\} & \text{otherwise.} \end{cases}$$

In the first case ($a_j = n$) the maximum number of samples allowed has already been collected, the second one ($a_j + N - j + 1 = n$) means that all remaining samples of the population must be accepted in order to reach a total of n samples selected. It is only in the last case that some freedom exists for decisions.

The case $d = \dim(\theta) = 1$ is considered in (Pronzato, 2001a). The optimal (closed-loop) solution is given by a backward recurrence equation. Using results on extreme value distributions, a simple open-loop solution is proved to be asymptotically optimal for $N \rightarrow \infty$ with n fixed (for measures μ absolutely continuous with respect to the Lebesgue measure and such that the associated distribution function is a von Mises function). This extends the results of Albright and Derman (1972) which concern the case $n = \lfloor \alpha N \rfloor$, $\alpha \in (0, 1)$. In the multidimensional case $d > 1$, in general, the optimal solution cannot be obtained in closed form. Open-loop feedback-optimal control is used in (Pronzato, 1999) and a heuristic one-step ahead decision rule in (Pronzato, 2001b), without any result on the asymptotic performance of these strategies.

The asymptotics considered in this paper will only concern the case $n = \lfloor \alpha N \rfloor$, $\alpha \in (0, 1)$, $N \rightarrow \infty$. The fact that n tends to infinity at the same speed as N means that we shall obtain asymptotic performances that are achieved μ -almost surely (a.s.), contrary to (Pronzato, 2001a) which concerns expected performances.

The different types of strategies to be considered (with finite or infinite horizon, stationary, non-adaptive, randomized) are defined in Section 2. The value $\phi(\xi_\alpha^*)$, with $\xi_\alpha^* \leq \mu/\alpha$ a ϕ -optimum bounded design measure, is shown to form an upper bound on the asymptotic performance of any strategy. A trivial

asymptotically optimal strategy $\mathcal{S}_{N,n}(\alpha)$ is proposed in Section 3: $\mathcal{S}_{N,n}(\alpha)$ is the truncated version (finite horizon) of a stationary strategy $\mathcal{S}(\xi_\alpha^*)$ which samples from ξ_α^* . We show in Section 4 that a slight modification of $\mathcal{S}(\xi_\alpha^*)$ leads to an asymptotically optimal strategy that does not require the prior knowledge of the measure μ . Illustrative examples are presented in Section 5. Section 6 gives some concluding remarks and suggests some extensions.

2 Strategies, performances and asymptotic optimality

2.1 Strategies

We shall consider *sequential* strategies for which at step k only X_1, \dots, X_k have been observed. When μ is known, $\Omega_k = (k, a_k, \mathbf{M}_{k-1, a_k}, X_k)$ then summarizes all the information necessary to make a decision about the acceptance of X_k , and the problem (3) corresponds to a discrete-time stochastic control problem, where k represents time, $(k, a_k, \mathbf{M}_{k-1, a_k}, X_k)$ and $u_k \in \mathcal{U}_k \subseteq \{0, 1\}$ respectively represent the state¹ and control at time k . A sequential strategy $\mathcal{S}_{N,n}$ is then defined by a sequence of mappings $(k, a, \mathbf{M}, X) \mapsto u \in \{0, 1\}$, $k = 1, \dots, N$. For instance, the optimal strategy defined by (5) is sequential.

Strategies are *non-sequential* when the selection of the n X_k 's is made after all the population X_1, \dots, X_N has been observed. A non-sequential strategy will be used to obtain an upper bound on the performance of sequential strategies.

¹ When μ is unknown, a case considered in Section 4, at step k an estimate $\hat{\mu}_k$ of μ (empirical version, parametric representation, etc.) based on X_1, \dots, X_k must also be used to decide about X_k ; the “state” Ω_k should then be extended to include this $\hat{\mu}_k$.

A sequential strategy \mathcal{S} will be called *stationary* when the mapping is from (\mathbf{M}, X) to $u \in \{0, 1\}$. Stationary strategies correspond to the case of infinite horizon ($N = \infty$) where we only require that $a_k/k \rightarrow \alpha$ when $k \rightarrow \infty$. We shall denote ξ_{k-1} the empirical measure of the a_k samples already accepted at step k , with $\mathbf{M}(\xi_{k-1})$ the current value of the (normalized) information matrix,

$$\mathbf{M}(\xi_{k-1}) = \frac{1}{a_k} \sum_{j=1}^{k-1} u_j f(X_j) f^\top(X_j),$$

and write

$$\phi_k = \phi(\xi_k) = \Phi[\mathbf{M}(\xi_k)].$$

Non-adaptive strategies are special cases of stationary strategies for which the decision at step k only depends on the value of X_k , and \mathcal{S} samples from some measure $\xi \leq \mu/\alpha$; we shall then write $\mathcal{S} = \mathcal{S}(\xi)$. We consider non-adaptive strategies as sequential, although the information carried by previous samples is not used.

The non-adaptive strategies $\mathcal{S}(\xi)$ that we shall consider have the property that $\xi = 0$ on \mathcal{X}_1 , $\xi = \mu/\alpha$ on \mathcal{X}_2 and $\xi = \beta\mu/\alpha$, $0 < \beta < 1$, on $\mathcal{X} \setminus (\mathcal{X}_1 \cup \mathcal{X}_2)$ for some subsets $\mathcal{X}_1, \mathcal{X}_2$ of \mathcal{X} . This means that $\mathcal{S}(\xi)$ should reject every X_k falling in \mathcal{X}_1 , accept all the samples in \mathcal{X}_2 but only a fraction β of those in \mathcal{X}_3 . Therefore, the strategy is completely defined when $\mathcal{X}_1 \cup \mathcal{X}_2 = \mathcal{X}$ but, when $\mathcal{X}_3 \neq \emptyset$, we must specify how the fraction β of the samples in \mathcal{X}_3 is selected. We shall then consider *randomized* strategies, which lead to a simple analysis: a sample in \mathcal{X}_3 will be accepted with probability β , by tossing a biased coin. The mapping $X \mapsto u \in \{0, 1\}$ is then a random function when $X \in \mathcal{X}_3$.

A stationary strategy, defined for infinite horizon, will be implemented when the horizon is finite through a truncation procedure, which will be shown to

preserve asymptotic optimality (when $n = \lfloor \alpha N \rfloor$, $\alpha \in (0, 1)$, $N \rightarrow \infty$).

2.2 Performances

For any strategy $\mathcal{S}_{N,n}$ used to solve (3), with $0 < n \leq N$, we shall denote $\Psi(\mathcal{S}_{N,n}) = \Phi(\mathbf{M}_{N,n})$, which measures the performance of $\mathcal{S}_{N,n}$ (for a particular realization of the sequence X_1, \dots, X_N).

Following the same line as in (Albright and Derman, 1972), which concerns the case $d = 1$, we can use as a benchmark the infeasible, but better-than-optimal, *non-sequential* strategy $\mathcal{S}_{N,n}^*$, obtained by selecting the n design points X_{k_1}, \dots, X_{k_n} that maximise $\Phi(\mathbf{M}_{N,n})$ after the N points X_1, \dots, X_N have been observed. $\mathcal{S}_{N,n}^*$ is thus a ϕ -optimum design algorithm that generates an exact n -point ϕ -optimum design in the finite design space $\{X_1, \dots, X_N\}$. Obviously, for any N , n and any strategy $\mathcal{S}_{N,n}$, sequential or not,

$$\Psi(\mathcal{S}_{N,n}) \leq \Psi(\mathcal{S}_{N,n}^*). \quad (6)$$

Also, $\mathcal{S}_{N,n}^*$ satisfies

$$\forall \alpha \in (0, 1), \limsup_{N \rightarrow \infty} \Psi(\mathcal{S}_{N, \lfloor \alpha N \rfloor}^*) \leq \phi_\alpha^*, \quad \mu\text{-a.s.} \quad (7)$$

with $\phi_\alpha^* = \phi(\xi_\alpha^*)$ and $\xi_\alpha^* \leq \mu/\alpha$ a ϕ -optimum constrained design measure: ξ_α^* maximises $\phi(\xi)$, $\xi \in \mathcal{D}(\mu, \alpha)$, with $\mathcal{D}(\mu, \alpha) \subset \Xi$ the set of admissible measures satisfying

$$\xi(dx) \leq \mu(dx)/\alpha.$$

When $\Phi(\cdot)$ is strictly concave ($\mathbf{H}_\Phi \mathbf{1}$), $\mathbf{M}(\xi_\alpha^*)$ is unique (although ξ_α^* is not necessarily unique). ($\mathbf{H}_\mu \mathbf{1}$) implies $\phi_\alpha^* \geq \phi(\mu) > -\infty$ and ($\mathbf{H}_\Phi \mathbf{3}$) implies

$\phi_\alpha^* \leq \phi(\mu/\alpha) < \infty$. Other properties of ξ_α^* will be presented in Section 3.1, see also Wynn (1982); Fedorov (1989); Fedorov and Hackl (1997); Sahm and Schwabe (2001); Pronzato (2004).

Following (6, 7), a strategy $\mathcal{S}_{N,n}$ will be called *asymptotically optimal* (for $n = \lfloor \alpha N \rfloor$, $\alpha \in (0, 1)$, $N \rightarrow \infty$) when $\lim_{N \rightarrow \infty} \Psi(\mathcal{S}_{N, \lfloor \alpha N \rfloor}) = \phi_\alpha^*$, μ -a.s.

Although sampling from a ϕ -optimum constrained measure ξ_α^* yields $\phi_k \rightarrow \phi_\alpha^*$, μ -a.s., $k \rightarrow \infty$ (we shall see in Section 3.2 how this can be implemented), it corresponds to a non-adaptive strategy that does not satisfy constraint (1). We show below how enforcing (1) can preserve asymptotic optimality.

For any $\alpha \in (0, 1)$ let \mathcal{S}_α be a stationary strategy such that $a_k/k \rightarrow \alpha$ and $\phi_k \rightarrow \phi_\alpha^*$, μ -a.s., $k \rightarrow \infty$. In particular, it may correspond to the non-adaptive strategy $\mathcal{S}(\xi_\alpha^*)$ which samples from a ϕ -optimum constrained measure $\xi_\alpha^* \leq \mu/\alpha$. To any such \mathcal{S}_α we associate the following truncation, which defines a sequential strategy $\mathcal{S}_{N,n}(\mathcal{S}_\alpha)$ satisfying the constraint $a_{N+1} = n$ for any finite $N \geq n$:

$$\mathcal{S}_{N,n}(\mathcal{S}_\alpha) : \left\{ \begin{array}{l} \text{if } a_k = n, \text{ reject } X_k; \\ \text{if } N - k + 1 = n - a_k, \text{ accept } X_k; \\ \text{otherwise, apply } \mathcal{S}_\alpha. \end{array} \right. \quad (8)$$

We show that $\mathcal{S}_{N,n}(\mathcal{S}_\alpha)$ is asymptotically optimal for $n = \lfloor \alpha N \rfloor$, $\alpha \in (0, 1)$, $N \rightarrow \infty$. Replace α by $\alpha + \epsilon$, with $\epsilon > 0$, and consider the following strategy,

simpler than $\mathcal{S}_{N,n}(\mathcal{S}_{\alpha+\epsilon})$:

$$\bar{\mathcal{S}}_{N,n}(\mathcal{S}_{\alpha+\epsilon}) : \left| \begin{array}{l} \text{if } a_k = n, \text{ reject } X_k; \\ \\ \text{otherwise, apply } \mathcal{S}_{\alpha+\epsilon}. \end{array} \right.$$

Let \bar{n} denote the number of samples accepted by $\bar{\mathcal{S}}_{N,n}(\mathcal{S}_{\alpha+\epsilon})$. We have $\bar{n} \leq n$, and, if $\mathbf{M}_{N,n}(\mathcal{S}_{\alpha+\epsilon})$ and $\bar{\mathbf{M}}_{N,n}(\mathcal{S}_{\alpha+\epsilon})$ denote the information matrices associated with both strategies, $\mathbf{M}_{N,n}(\mathcal{S}_{\alpha+\epsilon}) = \bar{\mathbf{M}}_{N,n}(\mathcal{S}_{\alpha+\epsilon})$ when $\bar{n} = n$ and $\mathbf{M}_{N,n}(\mathcal{S}_{\alpha+\epsilon}) = \bar{\mathbf{M}}_{N,n}(\mathcal{S}_{\alpha+\epsilon}) + (1/n) \sum_{j=1}^{n-\bar{n}} f(X'_j) f^\top(X'_j)$ when $\bar{n} < n$, where the X'_j 's denote the points selected by $\mathcal{S}_{N,n}(\mathcal{S}_{\alpha+\epsilon})$ and not $\bar{\mathcal{S}}_{N,n}(\mathcal{S}_{\alpha+\epsilon})$. Since $\Phi(\cdot)$ is increasing, we have

$$\Psi[\bar{\mathcal{S}}_{N,n}(\mathcal{S}_{\alpha+\epsilon})] \leq \Psi[\mathcal{S}_{N,n}(\mathcal{S}_{\alpha+\epsilon})]. \quad (9)$$

Let $n_{\alpha+\epsilon}$ denote the number of points accepted by $\mathcal{S}_{\alpha+\epsilon}$ among the N samples of the sequence (these are points that would have been accepted by $\bar{\mathcal{S}}_{N,n}(\mathcal{S}_{\alpha+\epsilon})$ if ignoring the constraint $a_k \leq n$). It satisfies $n_{\alpha+\epsilon}/N \rightarrow \alpha + \epsilon$, μ -a.s. as N tends to infinity. Take $n = \lfloor \alpha N \rfloor$ and let N tend to infinity. The probability that $n_{\alpha+\epsilon} \leq n$ infinitely often is zero when $\epsilon > 0$. Therefore, asymptotically, $\bar{\mathcal{S}}_{N, \lfloor \alpha N \rfloor}(\mathcal{S}_{\alpha+\epsilon})$ stops after $n = \lfloor \alpha N \rfloor$ samples from $\mathcal{S}_{\alpha+\epsilon}$ have been collected and

$$\Psi[\bar{\mathcal{S}}_{N, \lfloor \alpha N \rfloor}(\mathcal{S}_{\alpha+\epsilon})] \rightarrow \phi_{\alpha+\epsilon}^*, \quad \mu\text{-a.s.}, \quad N \rightarrow \infty.$$

Together with (9), it implies

$$\liminf_{N \rightarrow \infty} \Psi[\mathcal{S}_{N, \lfloor \alpha N \rfloor}(\mathcal{S}_{\alpha+\epsilon})] \geq \phi_{\alpha+\epsilon}^*, \quad \mu\text{-a.s.}, \quad N \rightarrow \infty.$$

On the other hand, from (6, 7), $\limsup_{N \rightarrow \infty} \Psi[\mathcal{S}_{N, \lfloor \alpha N \rfloor}(\mathcal{S}_{\alpha+\epsilon})] \leq \phi_{\alpha}^*$ μ -a.s.

We may then let ϵ tend to zero and use the continuity of ϕ_{α}^* with respect to

α , see Theorem 6, to obtain the following property.

Theorem 1 *For $\alpha \in (0, 1)$ let \mathcal{S}_α be a stationary strategy such that $a_k/k \rightarrow \alpha$ and $\phi_k \rightarrow \phi_\alpha^*$, μ -a.s., $k \rightarrow \infty$. The strategy $\mathcal{S}_{N,n}(\mathcal{S}_\alpha)$ defined by (8) is asymptotically optimal for $n = \lfloor \alpha N \rfloor$, $N \rightarrow \infty$.*

Section 3.1 will present some properties of ϕ -optimum constrained design measures, to be used in Section 3.2 to implement the stationary strategy $\mathcal{S}(\xi_\alpha^*)$. When $\mathcal{S}(\xi_\alpha^*)$ is plugged in (8), the resulting strategy $\mathcal{S}_{N,n}[\mathcal{S}(\xi_\alpha^*)]$ is asymptotically optimal from the previous theorem. Notice that from (6, 7) the existence of such an asymptotically optimal strategy implies

$$\lim_{N \rightarrow \infty} \Psi(\mathcal{S}_{N, \lfloor \alpha N \rfloor}^*) = \phi_\alpha^*, \quad \mu\text{-a.s.} \quad (10)$$

This property can be extended into the following (see Appendix B), which will be used in the proof of Theorem 9.

Lemma 2 *Let (α_k) be a sequence in $(0, 1)$, with $\lim_{k \rightarrow \infty} \alpha_k = \alpha \in (0, 1)$. Then, under $\mathbf{H}_\mu \mathbf{1}$, $\mathbf{H}_\Phi \mathbf{1}$ - $\mathbf{H}_\Phi \mathbf{3}$ and $\mathbf{H}_\Phi \mathbf{5}$,*

$$\lim_{N \rightarrow \infty} \Psi(\mathcal{S}_{N, \lfloor \alpha_N N \rfloor}^*) = \phi_\alpha^*, \quad \mu\text{-a.s.}$$

Whereas (10) and Lemma 2 only concern the limiting behavior of $\Psi(\mathcal{S}_{N,n}^*)$, the following result on $\mathbb{E}\{\Psi(\mathcal{S}_{N,n}^*)\}$ holds for any N under very general conditions. The proof is given in Appendix B.

Lemma 3 *For a concave criterion $\Phi(\cdot)$, the non sequential strategy $\mathcal{S}_{N,n}^*$ (that is, a ϕ -optimum algorithm for an exact design with n point in $\{X_1, \dots, X_N\}$)*

satisfies

$$\forall(n, N), 0 < n \leq N, \mathbf{E}\{\Psi(\mathcal{S}_{N,n}^*)\} \leq \phi(\xi_{n/N}^*), \quad (11)$$

with $\xi_{n/N}^*$ a ϕ -optimum constrained design measure in $\mathcal{D}(\mu, n/N)$.

H _{μ} 1 and **H _{Φ}** 3 imply $\phi(\xi) < \infty$ for any design measure $\xi \leq \mu/\alpha$, α given in $(0, 1)$, so that Lemma 3 implies that for any (n, N) , $0 < n \leq N$, and any strategy $\mathcal{S}_{N,n}$

$$\mathbf{E}\{\Psi(\mathcal{S}_{N,n})\} \leq \mathbf{E}\{\Psi(\mathcal{S}_{N,n}^*)\} \leq \phi(\xi_{n/N}^*) < \infty. \quad (12)$$

(When $\mathcal{S}_{N,n}$ involves randomized decisions, see e.g. Section 3.2, the first expectation in (12) is also with respect to them.)

It may be noticed that the upper bound $\phi(\xi_{n/N}^*)$ for the expected performance $\mathbf{E}\{\Psi(\mathcal{S}_{N,n})\}$ is not necessarily achievable *when n is finite, $N \rightarrow \infty$* and the strategy $\mathcal{S}_{N,n}$ is sequential, even in the case $d = \dim(\theta) = 1$, see Example 3 of (Pronzato, 2001a).

3 Constrained design measures and asymptotically optimum sequential strategies

3.1 Optimum constrained design measures

The main result (Wynn, 1982; Sahm and Schwabe, 2001), presented in the following theorem, states that when ϕ_α^* is a ϕ -optimum constrained measure, \mathcal{X} can be partitioned into three subsets $\mathcal{X}_{1,\alpha}^*$, $\mathcal{X}_{2,\alpha}^*$ and $\mathcal{X}_{3,\alpha}^* = \mathcal{X} \setminus (\mathcal{X}_{1,\alpha}^* \cup \mathcal{X}_{2,\alpha}^*)$,

with $\xi_\alpha^* = 0$ on $\mathcal{X}_{1,\alpha}^*$, $\xi_\alpha^* = \mu/\alpha$ on $\mathcal{X}_{2,\alpha}^*$ and the directional derivative $F(\xi_\alpha^*, x)$ (see Appendix A) constant on $\mathcal{X}_{3,\alpha}^*$.

Theorem 4 *The following statements are equivalent:*

- (i) ξ_α^* is a ϕ -optimum constrained design measure;
- (ii) there exists a number c such that $F(\xi_\alpha^*, x) \geq c$ for ξ_α^* -almost all x and $F(\xi_\alpha^*, x) \leq c$ for $(\mu - \alpha\xi_\alpha^*)$ -almost all x ;
- (iii) there exist two subsets $\mathcal{X}_{1,\alpha}^*$ and $\mathcal{X}_{2,\alpha}^*$ of \mathcal{X} such that
 - $\xi_\alpha^* = 0$ on $\mathcal{X}_{1,\alpha}^*$ and $\xi_\alpha^* = \mu/\alpha$ on $\mathcal{X}_{2,\alpha}^*$,
 - $\inf_{x \in \mathcal{X}_{2,\alpha}^*} F(\xi_\alpha^*, x) \geq c \geq \sup_{x \in \mathcal{X}_{1,\alpha}^*} F(\xi_\alpha^*, x)$,
 - $F(\xi_\alpha^*, x) = c$ on $\mathcal{X}_{3,\alpha}^* = \mathcal{X} \setminus (\mathcal{X}_{1,\alpha}^* \cup \mathcal{X}_{2,\alpha}^*)$.

The construction of the sets $\mathcal{X}_{1,\alpha}^*$ and $\mathcal{X}_{2,\alpha}^*$ is important in order to be able to sample from ξ_α^* ; also, we must precise the value of ξ_α^* on $\mathcal{X}_{3,\alpha}^* = \mathcal{X} \setminus (\mathcal{X}_{1,\alpha}^* \cup \mathcal{X}_{2,\alpha}^*)$. This is considered in the rest of this section.

For a given ξ , consider the random variable $F_\Phi(\xi, X_1)$ and let $\mathbb{F}_\xi(\cdot)$ denote the corresponding distribution function,

$$\mathbb{F}_\xi(s) = \mu\{x \mid F_\Phi(\xi, x) \leq s\}. \quad (13)$$

Define $c_\alpha(\xi)$ as

$$c_\alpha(\xi) = \min\{s \mid \mathbb{F}_\xi(s) \geq 1 - \alpha\} \quad (14)$$

and

$$\begin{aligned} \mathcal{X}_{1,\alpha}(\xi) &= \{x \mid F_\Phi(\xi, x) < c_\alpha(\xi)\}, \\ \mathcal{X}_{2,\alpha}(\xi) &= \{x \mid F_\Phi(\xi, x) > c_\alpha(\xi)\}, \\ \mathcal{X}_{3,\alpha}(\xi) &= \{x \mid F_\Phi(\xi, x) = c_\alpha(\xi)\}. \end{aligned} \quad (15)$$

We then obtain $\mathcal{X}_{j,\alpha}^* = \mathcal{X}_{j,\alpha}(\xi_\alpha^*)$, $j = 1, 2, 3$, and $c_\alpha(\xi_\alpha^*)$ is the constant c of Theorem 4.

Consider now the following transformation

$$T_{\Phi,\alpha} : \xi \in \Xi \rightarrow T_{\Phi,\alpha}(\xi) \in \mathcal{D}(\mu, \alpha), \quad T_{\Phi,\alpha}(\xi) = \begin{cases} \mu/\alpha \text{ on } \mathcal{X}_{2,\alpha}(\xi), \\ \frac{\alpha - \mu[\mathcal{X}_{2,\alpha}(\xi)]}{\mu[\mathcal{X}_{3,\alpha}(\xi)]} \mu/\alpha \text{ on } \mathcal{X}_{3,\alpha}(\xi), \\ 0 \text{ on } \mathcal{X}_{1,\alpha}(\xi). \end{cases}$$

Notice that $F_\Phi[\xi; T_{\Phi,\alpha}(\xi)] = \max_{\nu \in \mathcal{D}(\mu, \alpha)} F_\Phi(\xi; \nu)$, with $\mathcal{D}(\mu, \alpha)$ the set of admissible measures on \mathcal{X} satisfying $\nu \leq \mu/\alpha$ (indeed, $T_{\Phi,\alpha}(\xi)$ distributes its mass on \mathcal{X} where $F_\Phi(\xi, x)$ takes its highest values).

The next theorem complements Theorem 4 by a minimax formulation similar to the Kiefer-Wolfowitz (1960) Equivalence Theorem.

Theorem 5 *The following statements are equivalent:*

- (i) ξ_α^* is a ϕ -optimum constrained design measure;
- (ii) $F_\Phi[\xi_\alpha^*; T_{\Phi,\alpha}(\xi_\alpha^*)] = 0$;
- (iii) ξ_α^* minimises $F_\Phi[\xi; T_{\Phi,\alpha}(\xi)]$, $\xi \in \mathcal{D}(\mu, \alpha)$;
- (iv) ξ_α^* minimises $\max_{\nu \in \mathcal{D}(\mu, \alpha)} F_\Phi(\xi; \nu)$, $\xi \in \mathcal{D}(\mu, \alpha)$.

Proof. From Theorem 4, (i) gives $\phi[T_{\Phi,\alpha}(\xi_\alpha^*)] = \phi(\xi_\alpha^*)$ which implies (ii). Take any $\xi \in \Xi$ and denote ξ_α^* an optimum constrained design measure. From the definition of $T_{\Phi,\alpha}$,

$$\forall \nu \in \mathcal{D}(\mu, \alpha), \quad F_\Phi[\xi; T_{\Phi,\alpha}(\xi)] \geq F_\Phi(\xi; \nu) \tag{16}$$

and thus $F_\Phi[\xi; T_{\Phi, \alpha}(\xi)] \geq F_\Phi(\xi; \xi_\alpha^*)$. Concavity of Φ then gives

$$\forall \xi \in \Xi, \quad F_\Phi[\xi; T_{\Phi, \alpha}(\xi)] \geq \phi_\alpha^* - \phi(\xi). \quad (17)$$

Since $\phi_\alpha^* \geq \phi(\xi)$, $\xi \in \mathcal{D}(\mu, \alpha)$, (ii) implies (i). Equivalence between (i) and (iii) is obvious from (17); (16) implies the equivalence between (iii) and (iv). \square

We conclude this section by mentioning the following Lipschitz property, see (Pronzato, 2004), which shows that ϕ_α^* is continuous in α .

Theorem 6 *For any α, β in $(0, 1)$, the associated ϕ -optimum constrained design measures ξ_α^* and ξ_β^* satisfy*

$$|\phi_\alpha^* - \phi_\beta^*| \leq |\alpha - \beta| \max \left[\frac{|c_\alpha(\xi_\alpha^*)|}{\beta}, \frac{|c_\beta(\xi_\beta^*)|}{\alpha} \right],$$

with $c_\alpha(\xi)$ defined by (14).

3.2 An asymptotically optimum sequential strategy

The implementation of a non-adaptive strategy that samples from ξ_α^* is straightforward when μ has no atoms, that is, when for any $\Delta\mathcal{X}$ with $\mu(\Delta\mathcal{X}) > 0$ exists $\Delta\mathcal{X}' \subset \Delta\mathcal{X}$ such that $0 < \mu(\Delta\mathcal{X}') < \mu(\Delta\mathcal{X})$, with measures absolutely continuous w.r.t. the Lebesgue measure as a special case. Indeed, in that case there exists a ξ_α^* such that the set $\mathcal{X}_{3, \alpha}^*$ is empty, see (Wynn, 1982; Fedorov, 1989; Fedorov and Hackl, 1997), and a trivial implementation of $\mathcal{S}(\xi_\alpha^*)$ then consists in accepting the samples that fall into $\mathcal{X}_{2, \alpha}^*$ and rejecting the others.

We shall use the transformation $T_{\Phi, \alpha}$ of previous section to deal with the general situation, and sample from $T_{\Phi, \alpha}(\xi_\alpha^*)$ (which is equivalent to sampling

from ξ_α^* since $\phi[T_{\Phi,\alpha}(\xi_\alpha^*)] = \phi(\xi_\alpha^*)$, see the proof of Theorem 5). Since only the fraction $[\alpha - \mu(\mathcal{X}_{2,\alpha}^*)]/\mu(\mathcal{X}_{3,\alpha}^*)$ of the samples in $\mathcal{X}_{3,\alpha}^*$ must be accepted, we randomize the selection procedure and define

$$\mathcal{S}(\xi_\alpha^*) : \left\{ \begin{array}{l} \text{if } X_k \in \mathcal{X}_{2,\alpha}^* \text{ accept } X_k ; \\ \\ \text{if } X_k \in \mathcal{X}_{3,\alpha}^* \text{ accept } X_k \text{ with probability} \\ \\ P_{3,\alpha}^* = [\alpha - \mu(\mathcal{X}_{2,\alpha}^*)]/\mu(\mathcal{X}_{3,\alpha}^*) ; \\ \\ \text{otherwise reject } X_k . \end{array} \right. \quad (18)$$

From Theorem 1, when the non-adaptive strategy $\mathcal{S}(\xi_\alpha^*)$ is plugged in (8) the resulting $\mathcal{S}_{N,n}[\mathcal{S}(\xi_\alpha^*)]$ is asymptotically optimal for $n = \lfloor \alpha N \rfloor$, $N \rightarrow \infty$. However, it requires the construction of the sets $\mathcal{X}_{j,\alpha}^*$, $j = 1, 2, 3$, and thus of a ϕ -optimum constrained design measure $\xi_\alpha^* \leq \mu/\alpha$. We show in the next section that it is possible to avoid this construction while preserving asymptotic optimality.

Remark 7 *A deterministic version of (18) consists in accepting a X_k that falls into $\mathcal{X}_{3,\alpha}^*$ only when $a_k/k < \alpha$. One can then easily show that $a_k/k \rightarrow \alpha$ and $\phi_k \rightarrow \phi_\alpha^*$ as $k \rightarrow \infty$, μ -a.s. (however, the strategy is then non stationary, see Section 2.1).*

4 Sampling asymptotically from a constrained measure

Consider the following modification of the strategy $\mathcal{S}(\xi_\alpha^*)$ defined by (18): at step k , we simply substitute the sets $\mathcal{X}_{j,\alpha}(\xi_{k-1})$ for $\mathcal{X}_{j,\alpha}^*$, $j = 1, 2, 3$, with $\mathcal{X}_{j,\alpha}(\xi)$ defined by (15) and ξ_{k-1} the empirical measure defined by the a_k design points

already selected. We thus define the following adaptive strategy:

$$\begin{array}{l|l}
\mathcal{S}_\alpha(\mu) : & \begin{array}{l} \text{if } X_k \in \mathcal{X}_{2,\alpha}(\xi_{k-1}) \text{ accept } X_k ; \\ \\ \text{if } X_k \in \mathcal{X}_{3,\alpha}(\xi_{k-1}) \text{ accept } X_k \text{ with probability} \\ \\ P_k(\alpha) = \{\alpha - \mu[\mathcal{X}_{2,\alpha}(\xi_{k-1})]\} / \mu[\mathcal{X}_{3,\alpha}(\xi_{k-1})] ; \\ \\ \text{otherwise reject } X_k . \end{array}
\end{array} \tag{19}$$

Remark 8

1. Notice that the sets $\mathcal{X}_{j,\alpha}(\xi_{k-1})$ are obtained from the function $F_\Phi(\xi_{k-1}, x)$, that is, from $\mathbf{M}(\xi_{k-1})$. Hence, the (randomized) decision at step k only depends on X_k and $\mathbf{M}(\xi_{k-1})$ and $\mathcal{S}_\alpha(\mu)$ is stationary.

2. In practise, the first samples X_k are always accepted until $\mathbf{M}(\xi_{k-1})$ becomes nonsingular, but this initialization has no effect on the asymptotic behavior of $\mathcal{S}_\alpha(\mu)$. Also, for technical reasons, we can assume that X_k is always accepted when $a_k/k < \alpha/C$, with C an arbitrarily large constant. This has no practical importance for the asymptotic behavior of the strategy (since $a_k/k \rightarrow \alpha$ μ -a.s., see Theorem 9 below) and for that reason is not mentioned in the definition of $\mathcal{S}_\alpha(\mu)$. On the other hand, it implies $a_{k+1}/k > \alpha/C$ for any k which permits to guarantee that $\mathbb{E}\{|\phi_k|\} < \infty$, a property used in the proof of asymptotic optimality, see Theorem 9.

3. $\mathcal{S}_\alpha(\mu)$ takes a simpler form when $\mathbb{F}_\xi(s)$ given by (13) is continuous in s for any ξ (so that $\mathcal{X}_{3,\alpha}(\xi_{k-1})$ is always empty): we accept X_k when $\mathbb{F}_{\xi_{k-1}}(X_k) > 1 - \alpha$ and reject X_k otherwise, and the strategy is fully deterministic.

4. We observe the same behavior when $\mathcal{S}_\alpha(\mu)$ is modified as follows: any $X_k \in \mathcal{X}_{3,\alpha}(\xi_{k-1})$ is accepted if $a_k/k < \alpha$ and is rejected otherwise; see Remark 7.

However, the strategy is then non stationary and its asymptotic behavior is more difficult to analyse.

5. Obviously, the sequence (ϕ_k) generated by (19) is not monotonically increasing, since (i) the step-length $1/(1 + a_k)$ when X_k is accepted and ξ_{k-1} updated is predetermined and (ii) X_k is random. While (i) is standard in the construction of optimum designs, (ii) is less common and forms a specific feature of the context considered here. In order to eliminate the unboundedness case encountered in the dichotomous theorem of Wu and Wynn (1978), which is usually the main issue raised by (i), we introduce the assumption $\mathbf{H}_\mu \mathbf{3}$ on μ , see Appendix A.

The asymptotic behavior of (19) satisfies the following.

Theorem 9 Under $\mathbf{H}_\Phi \mathbf{1}$ - $\mathbf{H}_\Phi \mathbf{6}$ and $\mathbf{H}_\mu \mathbf{1}$ - $\mathbf{H}_\mu \mathbf{3}$ the empirical measure ξ_k defined by the points accepted by the stationary strategy $\mathcal{S}_\alpha(\mu)$ satisfies $a_k/k \rightarrow \alpha$, μ -a.s., and

$$\lim_{k \rightarrow \infty} \phi(\xi_k) = \phi(\xi_\alpha^*), \mu\text{-a.s.}, \quad (20)$$

as $k \rightarrow \infty$, with $\xi_\alpha^* \leq \mu/\alpha$ a ϕ -optimum constrained design measure.

The proof is given in Appendix B. When the strategy is truncated as indicated in (8), the resulting $\mathcal{S}_{N,n}[\mathcal{S}_\alpha(\mu)]$ is asymptotically optimal for $n = \lfloor \alpha N \rfloor$, $N \rightarrow \infty$, from Theorem 1. Another possible truncation consists in adapting α and using, at step k , the strategy $\mathcal{S}_{\alpha_k}(\mu)$, with

$$\alpha_k = \frac{n - a_k}{N - k + 1}, \quad (21)$$

and a_k defined by (4). $\mathcal{S}_{\alpha_k}(\mu)$ coincides with the one-step-ahead rule suggested in (Pronzato, 2001b): when $\alpha_k = 0$, $\mathcal{X}_{2,\alpha_k}(\xi_{k-1})$ is empty, $\mu[\mathcal{X}_{2,\alpha_k}(\xi_{k-1})] = 0$ and X_k is always rejected; when $\alpha_k \geq 1$, $\mathcal{X}_{2,\alpha_k}(\xi_{k-1}) = \mathcal{X}$ and X_k is always accepted.

When μ is unknown, we can use $\mathcal{S}_\alpha(\hat{\mu}_k)$, or $\mathcal{S}_{\alpha_k}(\hat{\mu}_k)$, at step k , with $\hat{\mu}_k$ the empirical version of μ (or a kernel estimate, or a parametric representation $\mu_{\hat{\beta}_k}$, with $\hat{\beta}_k$ estimated from X_1, \dots, X_k). The estimation of μ does not depend on the strategy that is used (which corresponds to a *separation property* in control theory), and Theorem 9 still holds: we can thus asymptotically sample from ξ_α^* , without constructing ξ_α^* beforehand and even without knowing μ in advance. Illustrative examples are presented in the next section.

Consider finally a nonlinear situation where the information matrix \mathbf{M} depends on the model parameters θ , so that local optimum design is based on $\mathbf{M}(\hat{\theta}^0)$ with $\hat{\theta}^0$ a nominal value for θ . When θ can be estimated on line, it is natural to replace $\hat{\theta}^0$ by an estimate, $\hat{\theta}^k$ at step k . When using least squares estimation in a nonlinear regression problem with independent observations $Y_k = \eta(X_k, \bar{\theta}) + \varepsilon_k$, where (ε_k) is a sequence of i.i.d. errors with zero mean and finite variance, consistency and asymptotic normality of $\hat{\theta}^k$ will hold under $\mathbf{H}_\mu \mathbf{3}$ (and additional conditions on higher order derivatives of $\eta(\theta, x)$ with respect to θ and their tail cross product, see Jennrich (1969)). The sampling strategy $\mathcal{S}_\alpha(\mu)$ will then ensure $\phi_k \rightarrow \phi[\xi_\alpha^*(\bar{\theta})]$, μ -a.s., $k \rightarrow \infty$, with $\xi_\alpha^*(\bar{\theta}) \leq \mu/\alpha$ a ϕ -optimum constrained design measure for the *true value* $\bar{\theta}$ of the model parameters. This is illustrated by Example 12 below.

5 Examples

We take $\Phi(\cdot) = \log \det(\cdot)$ in all the examples below. Note that all the conditions \mathbf{H}_Φ are then satisfied, see Appendix A.

Example 10 *We consider the quadratic regression model $\eta(\theta, X) = \theta_0 + \theta_1 X + \theta_2 X^2$. Let μ_c correspond to the normal distribution $\mathcal{N}(0, 1)$ and μ_d be the discrete measure supported at the points $\{-1, -1/2, 0, 1/2, 1\}$ with respective weights $(1/8, 1/4, 1/4, 1/4, 1/8)$. We take $\mu = 0.5\mu_c + 0.5\mu_d$ and $\alpha = 0.1$.*

Easy calculations show that the ϕ -optimum constrained measure ξ_α^ is equal to μ/α on $\mathcal{X}_\alpha^* = (-\infty, -a] \cup [a, \infty)$, with $a \simeq 1.5625$, and puts the rest of its weight, approximately 0.4091, at zero. Figure 1 presents a plot of the sensitivity function $d(\xi_\alpha^*, x) = F_\Phi(\xi_\alpha^*, x) + 3 = f^\top(x)\mathbf{M}^{-1}(\xi_\alpha^*)f(x)$ and illustrates the optimality of ξ_α^* . Figure 2, left, gives a histogram of the first 1000 samples accepted by $\mathcal{S}_\alpha(\hat{\mu}_k)$, with $\hat{\mu}_k$ the empirical measure of the X_k 's. The right part of the figure presents ϕ_k as a function of k (log scale), with the optimal value $\phi_{0.1}^*$ indicated by the dashed line. (Note the fast convergence of ϕ_k , the large value of N being required only to have enough points for the histogram plot.)*

Example 11 *We consider again the case of a measure μ having both discrete and continuous components. The response of the regression model is $\eta(\theta, X) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$ with $\theta = (\theta_0, \theta_1, \theta_2)$ and X is two-dimensional, $X = (x_1, x_2)$. The continuous component μ_c corresponds to the normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I}_2)$, with \mathbf{I}_2 the 2-dimensional identity matrix, the discrete component μ_d puts weight $1/4$ at each one of the points $(\pm 1, \pm 1)$. Define $\mathcal{B}(a) = \{\mathbf{x} / \|\mathbf{x}\| > a\}$. When $\mu = (1/2)(\mu_c + \mu_d)$ the results are as follows.*

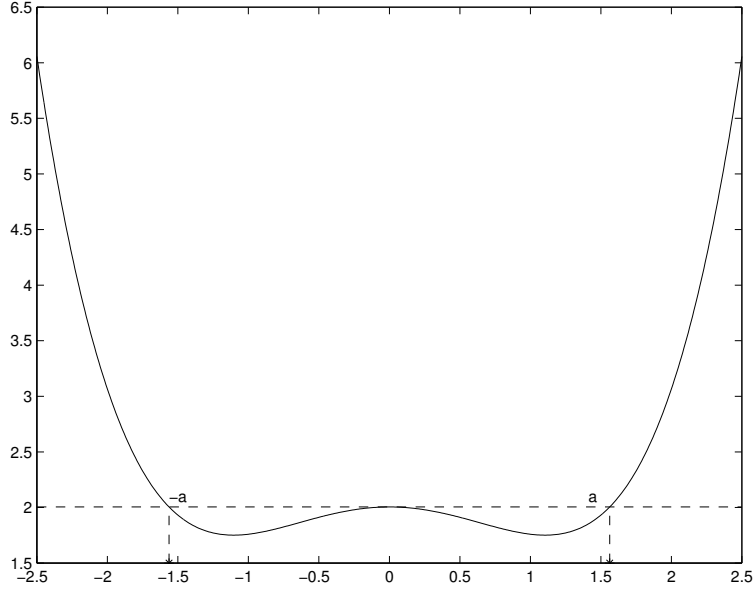


Fig. 1. Sensitivity function $d(\xi_\alpha^*, x)$ for the optimal constrained measure ξ_α^* in Example 10.

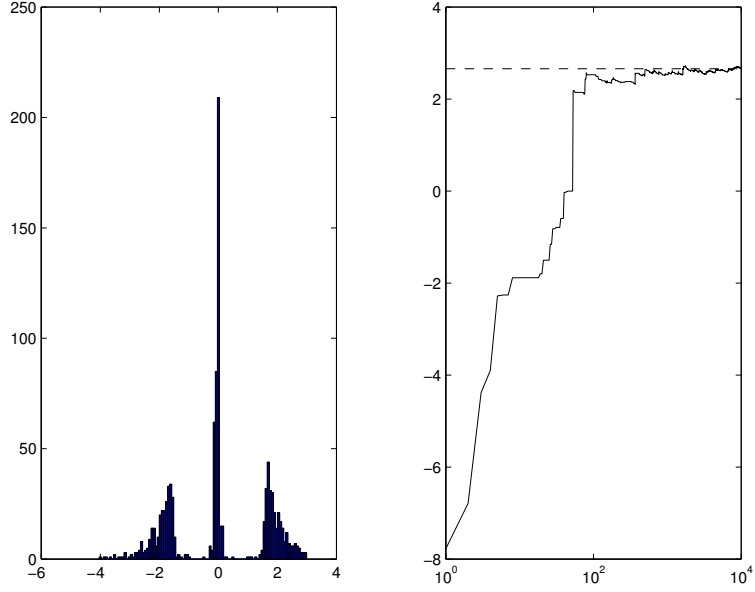


Fig. 2. Left: histogram of the 1000 first samples X_k accepted by $\mathcal{S}_\alpha(\hat{\mu}_k)$ in Example 10. Right: ϕ_k as a function of k (log scale); the value of $\phi_{0.1}^*$ corresponds to the dashed line.

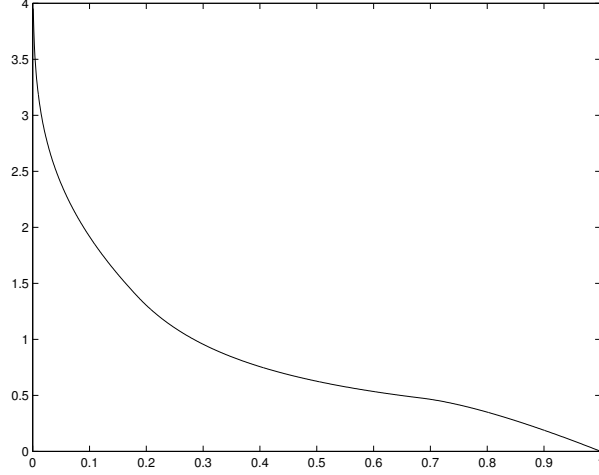


Fig. 3. ϕ_α^* as a function of α in Example 11.

For $0 < \alpha \leq 1/(2e)$, with $e = \exp(1)$, ξ_α^* coincides with $\mu_c/(2\alpha)$ on $\mathcal{B}(a_\alpha)$ with $a_\alpha = \sqrt{2 \log[1/(2\alpha)]} \geq \sqrt{2}$; $\phi_\alpha^* = 2 \log[1 - \log(2\alpha)]$.

For $1/(2e) < \alpha \leq 1/(2e) + 1/2$, $\xi_\alpha^* = \mu/\alpha$ on $\mathcal{B}(\sqrt{2})$ and $\xi_\alpha^* = 2[\alpha - 1/(2e)]\mu/\alpha$ on the four points $(\pm 1, \pm 1)$; $\phi_\alpha^* = 2 \log[1 + 1/(2e\alpha)]$.

For $1/(2e) + 1/2 < \alpha \leq 1$, $\xi_\alpha^* = \mu/\alpha$ on $\mathcal{B}(b_\alpha)$ with $b_\alpha = \sqrt{2 \log[1/(2\alpha - 1)]} < \sqrt{2}$; $\phi_\alpha^* = 2 \log\{1 - [(2\alpha - 1) \log(2\alpha - 1)]/(2\alpha)\}$.

Figure 3 presents ϕ_α^* as a function of α (although it does not appear clearly from the figure, ϕ_α^* tends to infinity when α tends to zero). Here $c_\alpha(\xi_\alpha)$ is a continuous function of α , so that ϕ_α^* is differentiable with respect to α , see Pronzato (2004).

We take $N = 10\,000$, $n = 2000$ and use $\mathcal{S}_{\alpha_k}(\hat{\mu}_k)$, see (19,21), with $\hat{\mu}_k$ the empirical measure of the X_k 's. (Note that $\alpha = 0.2$ is related to the intermediate case $1/(2e) < \alpha \leq 1/(2e) + 1/2$.) In the simulation presented $\mathbf{M}(\xi_k)$ has full rank after the first three samples X_1, X_2, X_3 have been accepted, and Figure 4 (top) presents a histogram of the remaining 1997 points accepted by $\mathcal{S}_{\alpha_k}(\hat{\mu}_k)$.

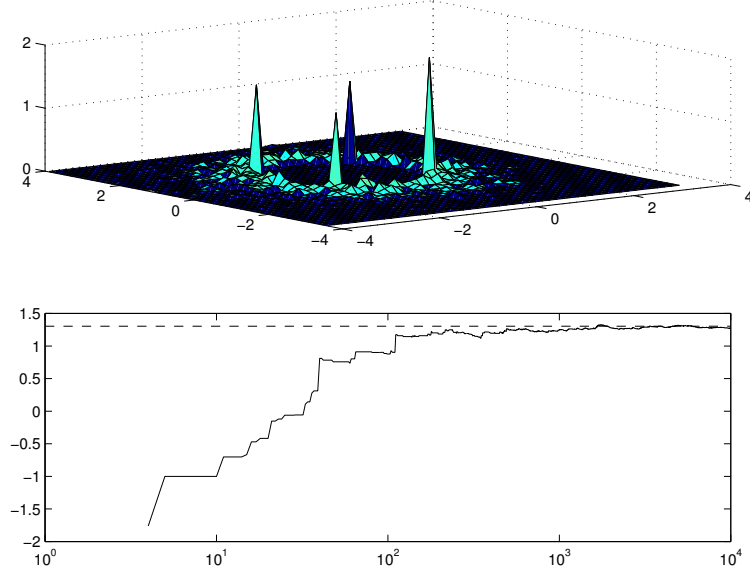


Fig. 4. Top: histogram of the 2000 points accepted by $\mathcal{S}_{\alpha_k}(\hat{\mu}_k)$ in a sequence of 10 000 points; bottom: ϕ_k as a function of k (log scale), the value of $\phi_{0.2}^*$ corresponds to the dashed line (Example 11).

Figure 4 (bottom) gives ϕ_k as a function of $k = 4, \dots, 2000$ (log scale), with the optimal value $\phi_{0.2}^* \simeq 1.3043$ indicated by the dashed line. (Again, notice the fast convergence of ϕ_k , the large value of N being required only to have enough points for the histogram plot.)

Example 12 We consider now the nonlinear regression model used in (Box and Lucas, 1959), where

$$\mathbb{E}\{Y_k|X_k = x; \theta\} = \eta(\theta, x) = \frac{\theta_1}{\theta_1 - \theta_2} [\exp(-\theta_2 x) - \exp(-\theta_1 x)].$$

We estimate $\theta = (\theta_1, \theta_2)$ by LS, and use at step k the information matrix corresponding to the current estimated value $\hat{\theta}^k$. The numerical values used to generate the observations Y_k correspond to $\bar{\theta} = (0.7, 0.2)$ and the standard deviation of the measurement errors is 0.2. (With this value for $\bar{\theta}$, the D-optimal experiment corresponds to performing the same number of observations at $x \simeq 1.23$ and 6.86.)

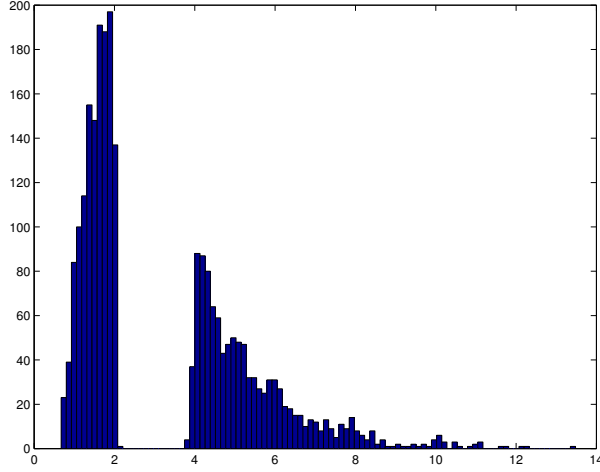


Fig. 5. Histogram of the 2,500 first samples X_k accepted by $\mathcal{S}_\alpha(\hat{\mu}_k)$ in Example 12. We assume that the experimental variables X_k have a lognormal distribution: $\log(X_k)$ is normally distributed $\mathcal{N}(1, 0.25)$. Figure 5 gives a histogram of the first 2,500 points accepted by $\mathcal{S}_\alpha(\hat{\mu}_k)$, with $\hat{\mu}_k$ the empirical measure of the X_k 's, when $\alpha = 0.5$. Easy calculations show that $\xi_\alpha^*(\bar{\theta}) = 2\mu$ on $[0, a] \cup [b, \infty)$, with $a \simeq 1.996$ and $b \simeq 3.922$, and $\phi[\xi_\alpha^*(\bar{\theta})] \simeq -2.143$.

6 Concluding remarks and further developments

Possible extensions of these results, that will be the subject of future work, include the following situations.

First, there are cases where the design variables are not directly observed: an example is when one observes covariates Z_k and the conditional probability measure $\mu(\cdot|Z_k)$ for the experimental conditions X_k is known for any k . A sequential selection strategy for this problem might reveal useful in phase-I clinical trials, where covariates Z_k such as the size, weight and age of volunteers could be used for their selection, in order to build a model of the tolerance dose as a function of (unobserved) pharmacokinetic/pharmacodynamic variables

X_k .

Second, applications to parameter estimation in dynamical systems, nonlinear in particular, call for an extension to correlated design variables X_k . A simple example is when $X_k = (U_k, U_{k-1}, \dots, U_{k-m})$, with (U_i) a random input sequence for the system. Note, however, that when the model contains an autoregressive part, that is, when $X_k = (U_k, U_{k-1}, \dots, U_{k-m}, Y_{k-1}, \dots, Y_{k-l})$, the decision not to observe Y_k implies that l future experimental conditions are unknown, which makes the problem much different from the one we considered here and will require specific developments.

In the one dimensional case $d = \dim(\theta) = 1$, asymptotic optimality of a strategy similar to $\mathcal{S}_{\alpha_k}(\mu)$, see (19, 21), is proved in (Pronzato, 2001a) for $N \rightarrow \infty$ with n fixed, provided the distribution function of X is a von Mises function (see, e.g., Embrechts et al. 1997, p. 138), with a tail decreasing faster than any power law. Extending Theorem 1 to the situation where n is fixed but $d > 1$ remains an open issue. Note in particular that in this case, although (12) gives an upper bound on the expected performance, the optimal performance achievable by a sequential strategy is unknown (not to speak about the optimal strategy itself). Also, all intermediate situations, between n fixed and $n = \lfloor \alpha N \rfloor$, such as $n = \lfloor \log N \rfloor$, or $\lfloor N^\beta \rfloor$ with $\beta < 1$, etc., are of interest. A possible application concerns the construction of optimum design algorithms. Indeed, classical algorithms for the determination of ξ^* (unconstrained) that maximizes $\phi(\cdot)$ rely on the determination at iteration k of a design point X_k that maximizes $F_\Phi(\xi_{k-1}, x)$ with respect to $x \in \mathcal{X}$, with ξ_{k-1} the current design measure. This (global) maximisation problem may prove cumbersome, especially if \mathcal{X} is high dimensional, so that it is sometimes recommended to accept any X_k such that $F_\Phi(\xi_{k-1}, X_k) > \delta$, with δ some small positive number,

see Fedorov and Hackl (1997), p. 49. A consequence of the results presented above is that generating candidates X_k randomly with some suitably chosen measure μ , with an acceptance rule such as (19), will ensure the convergence of ξ_k to ξ^* , provided α tends to zero at a proper speed. Whether or not efficient algorithms can be obtained in this way remains an open issue.

A Appendix (Assumptions and notations)

H_Φ1: Φ is strictly concave and $\Phi(\mathbf{M}) > -\infty$ for non singular \mathbf{M} .

H_Φ2: Φ is linearly differentiable; that is, the directional derivative

$$\mathcal{F}_\Phi(\mathbf{M}_1, \mathbf{M}_2) = \lim_{\epsilon \rightarrow 0^+} \{\Phi[(1 - \epsilon)\mathbf{M}_1 + \epsilon\mathbf{M}_2] - \Phi(\mathbf{M}_1)\} / \epsilon$$

satisfies $F_\Phi(\xi_1; \xi_2) = \mathcal{F}_\Phi[\mathbf{M}(\xi_1), \mathbf{M}(\xi_2)] = \int_{\mathcal{X}} F_\Phi(\xi_1, x) \xi_2(dx)$ for any ξ_1, ξ_2 with $\phi(\xi_1) = \Phi[\mathbf{M}(\xi_1)] > -\infty$, where $F_\Phi(\xi, x) = F_\Phi(\xi; \delta_x)$ and δ_x is the Dirac measure supported at x .

H_Φ3: Φ is increasing: $\mathbf{M}_2 - \mathbf{M}_1$ non negative definite implies $\Phi(\mathbf{M}_2) \geq \Phi(\mathbf{M}_1)$.

H_Φ4: $\lambda_{\min}(\mathbf{M}) \geq l > 0$ implies $\mathcal{F}_\Phi(\mathbf{M}, ff^\top) \geq -\delta_1(l)$ for any f , for some $\delta_1(l) < \infty$.

H_Φ5: there exist a function $g_1(\cdot)$ from \mathbb{R}^+ to \mathbb{R} and a function $g_2(\cdot)$ from \mathbb{R}^+ to \mathbb{R}^+ such that $\lim_{a \rightarrow 1^+} g_1(a) = 0$, $\lim_{a \rightarrow 1^+} g_2(a) = 1$ and $\Phi(a\mathbf{M}) \leq g_1(a) + g_2(a) \Phi(\mathbf{M})$ for any non-negative definite \mathbf{M} and any $a \geq 1$. Note that this assumption is satisfied for homogeneous criteria with $g_1 = 0$ and g_2 the identity. Also, if one wishes that maximizing $\Phi(\mathbf{M})$ be equivalent to maximiz-

ing $\Phi(a\mathbf{M})$ (which is often implicitly assumed when approximate designs are used), the assumption $\Phi(a\mathbf{M}) = g_1(a) + g_2(a)\Phi(\mathbf{M})$ comes naturally.

H_Φ6: Φ is two times differentiable. We denote by $\nabla_{\Phi}^2(\mathbf{M}_1, \mathbf{M}_2)$ the second order directional derivative

$$\nabla_{\Phi}^2(\mathbf{M}_1, \mathbf{M}_2) = \frac{\partial^2 \Phi[(1-\gamma)\mathbf{M}_1 + \gamma\mathbf{M}_2]}{\partial \gamma^2} \Big|_{\gamma=0^+}.$$

It satisfies, for any $\|f\| \leq R$ and \mathbf{M} such that $\lambda_{\min}(\mathbf{M}) \geq l > 0$: $\nabla_{\Phi}^2(\mathbf{M}, ff^{\top}) \geq -\delta_0(l) - \delta_2(l)R^4$, for some $\delta_0(l) < \infty$ and $\delta_2(l) < \infty$.

H_μ1: $-\infty < \phi(\mu) < \infty$.

H_μ2: $\mu_4 = \int_{\mathcal{X}} \|f(x)\|^4 \mu(dx) < \infty$. Note that Chebyshev's inequality implies $\text{Prob}\{\|f(X_1)\| \geq R\} \leq \mu_4/R^4$.

H_μ3: For $\alpha \in (0, 1)$ the proportion of interest, there exists $\underline{l} > 0$ and $\epsilon \in (0, \alpha)$ such that for any design measure $\xi \leq \mu/(\alpha - \epsilon)$, $\lambda_{\min}[\mathbf{M}(\xi)] > \underline{l}$.

Discussion of the assumptions.

In the case of D -optimality, where $\Phi(\mathbf{M}) = \log \det \mathbf{M}$, we have $F_{\Phi}(\xi, x) = f^{\top}(x)\mathbf{M}^{-1}(\xi)f(x) - d$ with $d = \dim(\theta)$. **H_Φ1-H_Φ4** are satisfied, with $\delta_1(l) = d$ in **H_Φ4**. **H_Φ5** is satisfied with $g_1(a) = d \log a$ and $g_2(a) = 1$. Finally, $\nabla_{\Phi}^2(\mathbf{M}_1, \mathbf{M}_2) = -\text{trace}\{[\mathbf{M}_1^{-1}(\mathbf{M}_2 - \mathbf{M}_1)]^2\}$ so that

$$\nabla_{\Phi}^2(\mathbf{M}, ff^{\top}) = -(f^{\top}\mathbf{M}^{-1}f)^2 + 2(f^{\top}\mathbf{M}^{-1}f) - d$$

and **H_Φ6** is satisfied with $\delta_0 = d$, $\delta_2(l) = 1/l^2$.

Take now $\Phi(\mathbf{M}) = -\text{trace}(\mathbf{A}\mathbf{M}^{-p})$, with p a positive integer and \mathbf{A} a positive definite matrix (Kiefer's Φ_p -class of optimality criteria). We can as-

sume that $\mathbf{A} = \mathbf{I}_d$, the d -dimensional identity matrix, by a linear transformation on the set $\mathbf{M}(\Xi) = \{\mathbf{M}(\xi), \xi \in \Xi\}$. We have $\mathcal{F}_\Phi(\mathbf{M}_1, \mathbf{M}_2) = p \text{trace}[\mathbf{M}_1^{-(p+1)}(\mathbf{M}_2 - \mathbf{M}_1)]$. $\mathbf{H}_\Phi \mathbf{1}$ - $\mathbf{H}_\Phi \mathbf{4}$ are satisfied, with $\delta_1(l) = p/l^p$ in $\mathbf{H}_\Phi \mathbf{4}$. $\mathbf{H}_\Phi \mathbf{5}$ is satisfied with $g_1(a) = 0$ and $g_2(a) = a^{-d}$. Finally, $\nabla_\Phi^2(\mathbf{M}_1, \mathbf{M}_2) = -p \sum_{a+b=p+2, a, b \geq 1} \text{trace}[\mathbf{M}_1^{-a}(\mathbf{M}_2 - \mathbf{M}_1)\mathbf{M}_1^{-b}(\mathbf{M}_2 - \mathbf{M}_1)]$, see Wu and Wynn (1978), and easy calculations give $\nabla_\Phi^2(\mathbf{M}, ff^\top) = -p \sum_{a+b=p+2, a, b \geq 1} [(f^\top \mathbf{M}^{-b} f) (f^\top \mathbf{M}^{-a} f)] + 2p(p+1)f^\top \mathbf{M}^{-(p+1)} f - p(p+1) \text{trace}[\mathbf{M}^{-p}]$. Therefore, $\mathbf{H}_\Phi \mathbf{6}$ is satisfied with $\delta_0(l) = p(p+1)/l^p$, $\delta_2(l) = p/l^{p+2}$.

$\mathbf{H}_\mu \mathbf{1}$ - $\mathbf{H}_\mu \mathbf{3}$ are satisfied for instance when: (i) the functions $f(x)$ are linearly independent on any open subset of \mathcal{X} with finite Lebesgue measure, (ii) μ has a component μ_c absolutely continuous with respect to the Lebesgue measure, with density φ , and a finite number of discrete components, (iii) the mass of μ_c is larger than $1 - \alpha + \epsilon$ (that is, the mass of the discrete components is less than $\alpha - \epsilon$), and (iv) \mathcal{X} is compact or $\varphi(x)$ is exponentially decreasing when $\|x\| \rightarrow \infty$.

B Appendix (Proofs)

Proof of Lemma 2. From $\mathbf{H}_\Phi \mathbf{3}$, for $n_1 \leq n_2 \leq N$, $\Psi(\mathcal{S}_{N, n_1}^*) = \Phi(\mathbf{M}_{N, n_1}^*) \leq \Phi[(n_2/n_1)\mathbf{M}_{N, n_2}^*]$, with $\mathbf{M}_{N, n}^*$ the *normalized* information matrix associated with the n points selected by $\mathcal{S}_{N, n}^*$, see (2). Since $\alpha_N \rightarrow \alpha$, for any ϵ such that $0 < \epsilon < \min(\alpha, 1 - \alpha)$ and for N large enough, $\alpha - \epsilon < \alpha_N < \alpha + \epsilon$. This implies that there exists N_0 such that, for any $N > N_0$,

$$\Phi(\mathbf{M}_{N, \lfloor \alpha_N N \rfloor}^*) \leq \Phi\left(\frac{\alpha + \epsilon}{\alpha - \epsilon} \mathbf{M}_{N, \lfloor (\alpha + \epsilon) N \rfloor}^*\right)$$

and

$$\Phi(\mathbf{M}_{N, \lfloor (\alpha - \epsilon)N \rfloor}^*) \leq \Phi\left(\frac{\alpha + \epsilon}{\alpha - \epsilon} \mathbf{M}_{N, \lfloor \alpha_N N \rfloor}^*\right).$$

Denote $a_\epsilon = (\alpha + \epsilon)/(\alpha - \epsilon)$, notice that $a_\epsilon > 1$ for $\alpha > 0$ and $a_\epsilon \rightarrow 1$ when $\epsilon \rightarrow 0$. **H_Φ5** then gives

$$\frac{\Phi(\mathbf{M}_{N, \lfloor (\alpha - \epsilon)N \rfloor}^*) - g_1(a_\epsilon)}{g_2(a_\epsilon)} \leq \Phi(\mathbf{M}_{N, \lfloor \alpha_N N \rfloor}^*) \leq g_2(a_\epsilon) \Phi(\mathbf{M}_{N, \lfloor (\alpha + \epsilon)N \rfloor}^*) + g_1(a_\epsilon)$$

and thus, from (10),

$$\begin{aligned} \limsup_{N \rightarrow \infty} \Phi(\mathbf{M}_{N, \lfloor \alpha_N N \rfloor}^*) &\leq g_2(a_\epsilon) \phi_{\alpha + \epsilon}^* + g_1(a_\epsilon) \\ \frac{\phi_{\alpha - \epsilon}^* - g_1(a_\epsilon)}{g_2(a_\epsilon)} &\leq \liminf_{N \rightarrow \infty} \Phi(\mathbf{M}_{N, \lfloor \alpha_N N \rfloor}^*), \end{aligned}$$

μ -a.s. The continuity of ϕ_α^* with respect to α (Theorem 6) and $\lim_{a \rightarrow 1^+} g_1(a) = 0$, $\lim_{a \rightarrow 1^+} g_2(a) = 1$ give $\lim_{N \rightarrow \infty} \Phi(\mathbf{M}_{N, \lfloor \alpha_N N \rfloor}^*) = \phi_\alpha^*$, μ -a.s. \square

Proof of Lemma 3. The strategy $\mathcal{S}_{N,n}^*$ satisfies $\Psi(\mathcal{S}_{N,n}^*) = \Phi(\mathbf{M}_{N,n}^*)$ with

$$\mathbf{M}_{N,n}^* = \frac{1}{n} \sum_{i=1}^N f(X_i) f^\top(X_i) J_n[X_j(j = 1, \dots, N, j \neq i), X_i],$$

where $J_n \in \{0, 1\}$ defines the decision function (the decision to accept X_i depends on the rest of the sequence). Therefore,

$$\mathbf{E}\{\mathbf{M}_{N,n}^*\} = \frac{1}{n} \int_{\mathcal{X}^N} \left\{ \sum_{i=1}^N f(x_i) f^\top(x_i) J_n[x_j(j = 1, \dots, N, j \neq i), x_i] \right\} \prod_{j=1}^N \mu(dx_j).$$

Using the fact that decisions are invariant by any permutation of the sequence, we get

$$\mathbf{E}\{\mathbf{M}_{N,n}^*\} = \int_{\mathcal{X}} f(x) f^\top(x) \xi_{N,n}^*(dx) = \mathbf{M}(\xi_{N,n}^*)$$

where

$$\xi_{N,n}^*(dx) = \frac{N}{n} \left(\int_{\mathcal{X}^{N-1}} J_n[x_j(j = 1, \dots, N-1), x] \prod_{j=1}^{N-1} \mu(dx_j) \right) \mu(dx).$$

Since $J_n \in \{0, 1\}$, $\xi_{N,n}^*(dx)$ satisfies $\xi_{N,n}^*(dx) \leq (N/n)\mu(dx)$. This implies $\Phi[\mathbf{M}(\xi_{N,n}^*)] \leq \phi(\xi_{n/N}^*)$. Concavity of $\Phi(\cdot)$ finally gives

$$\mathbb{E}\{\Phi(\mathbf{M}_{N,n}^*)\} \leq \Phi[\mathbb{E}\{\mathbf{M}_{N,n}^*\}] \leq \phi(\xi_{n/N}^*)$$

which concludes the proof. \square

Proof of Theorem 9.

Since X_k is accepted with probability α , $a_k/k \rightarrow \alpha$ μ -a.s. as $k \rightarrow \infty$.

The rest of the proof is decomposed in three steps. In **(i)**, we construct a lower bound on $\mathbb{E}\{\phi_k | X_1^{k-1}\}$, where X_1^{k-1} denotes the collection X_1, \dots, X_{k-1} . In **(ii)** we show that $\limsup_{k \rightarrow \infty} \phi_k = \phi_\alpha^*$, μ -a.s. Finally, in **(iii)** we show that $\liminf_{k \rightarrow \infty} \phi_k = \phi_\alpha^*$, μ -a.s., by an approach similar to Doob's upcrossing Lemma, see Williams (1991, p. 108).

(i) From **H_Φ2**, **H_Φ6**, criterion value ϕ_k associated with the empirical design measure ξ_k generated by $\mathcal{S}_\alpha(\mu)$ satisfies the recurrence

$$\begin{aligned} \phi_k = \phi_{k-1} + & \left[\frac{1}{1+a_k} F_\Phi(\xi_{k-1}, X_k) + \frac{1}{2(1+a_k)^2} H_\Phi(\xi_{k-1}, X_k, \gamma_k) \right] \\ & \times \left[\mathcal{I}_{\mathcal{X}_{2,k}}(X_k) + \mathcal{I}_{[0, P_k(\alpha)]}(Z) \mathcal{I}_{\mathcal{X}_{3,k}}(X_k) \right], \end{aligned} \quad (\text{B.1})$$

where a_k is defined by (4), Z is a random variable uniformly distributed in $[0, 1]$, $\mathcal{I}_{\mathcal{A}}(\cdot)$ denotes the indicator function of the set \mathcal{A} , and

$$H_\Phi(\xi_{k-1}, X_k, \gamma_k) = \nabla_\Phi^2[(1-\gamma_k)\mathbf{M}(\xi_{k-1}) + \gamma_k f(X_k)f^\top(X_k), f(X_k)f^\top(X_k)]$$

for some $\gamma_k \in [0, 1/(1+a_k)]$.

Since $a_k/k \rightarrow \alpha$ μ -a.s., from **H_μ3** there exists K_0 (μ -a.s.) such that for any $k > K_0$, $\lambda_{\min}[\mathbf{M}(\xi_{k-1})] > \underline{l}/2$. Since $\lambda_{\min}[(1-\gamma_k)\mathbf{M}(\xi_{k-1}) + \gamma_k f(X_k)f^\top(X_k)] \geq$

$[1 - 1/(1 + a_k)]\lambda_{\min}[\mathbf{M}(\xi_{k-1})]$, there exists K_1 (μ -a.s.) such that for $k > K_1$, $\lambda_{\min}[(1 - \gamma_k)\mathbf{M}(\xi_{k-1}) + \gamma_k f(X_k) f^\top(X_k)] > \underline{L}/4$. From **H_Φ6**, this implies $H_\Phi(\xi_{k-1}, X_k, \gamma_k) > -\delta_0(\underline{L}/4) - \delta_2(\underline{L}/4) \|f(X_k)\|^4$. We can now compute a lower bound on $\mathbb{E}\{\phi_k | X_1^{k-1}\}$. Notice that

$$\begin{aligned} \mathbb{E}\{F_\Phi(\xi_{k-1}, X_k) [\mathcal{I}_{\mathcal{X}_{2,k}}(X_k) + \mathcal{I}_{[0, P_k(\alpha)]}(Z)\mathcal{I}_{\mathcal{X}_{3,k}}(X_k)] | X_1^{k-1}\} = \\ \alpha F_\Phi[\xi_{k-1}; T_{\Phi, \alpha}(\xi_{k-1})] \end{aligned}$$

which gives for $k > K_1$,

$$\mathbb{E}\{\phi_k | X_1^{k-1}\} \geq \phi_{k-1} + \frac{\alpha}{1 + a_k} F_\Phi[\xi_{k-1}; T_{\Phi, \alpha}(\xi_{k-1})] - \alpha \frac{\delta_0(\underline{L}/4) + \delta_2(\underline{L}/4)\mu_4}{2(1 + a_k)^2}$$

with $\mu_4 = \int_{\mathcal{X}} \|f(x)\|^4 \mu(dx)$, and $\mu_4 < \infty$ from **H_μ2**. Using (17), we obtain

$$\mathbb{E}\{\phi_k | X_1^{k-1}\} \geq \phi_{k-1} + \frac{\alpha}{1 + a_k} (\phi_\alpha^* - \phi_{k-1}) - \frac{A}{(1 + a_k)^2} \quad (\text{B.2})$$

for $k > K_1$, where $A = \alpha [\delta_0(\underline{L}/4) + \delta_2(\underline{L}/4)\mu_4]/2$.

(ii) Assume that $\limsup \phi_k < \phi_\alpha^* - \delta$ for some $\delta > 0$, (B.2) gives

$$\mathbb{E}\{\phi_k | X_1^{k-1}\} > \phi_{k-1} + \frac{\alpha\delta}{2(1 + a_k)} \quad (\text{B.3})$$

for k larger than some K_2 . Since $\phi_k < \Psi(\mathcal{S}_{k, a_{k+1}}^*)$ and $a_{k+1}/k > \alpha/C$ for any k , see Remark 8.2, we have $-\infty \leq \mathbb{E}\{\phi_k\} < \mathbb{E}\{\Psi(\mathcal{S}_{k, a_{k+1}}^*)\} < \phi(\xi_{a_{k+1}/k}^*) < \phi(\xi_{\alpha/C}^*) < \infty$ (see Lemma 3, **H_μ1**, **H_Φ3**) and $\mathbb{E}\{\phi_k\}$ is well defined. Also, for $k > K_2$, (B.3) gives $\mathbb{E}\{\phi_k\} > \mathbb{E}\{\phi_{K_2}\} > -\infty$ and thus $\sup_k \mathbb{E}\{|\phi_k|\} < \infty$. The martingale convergence theorem then says that ϕ_k converges μ -a.s. to a finite limit, which contradicts (B.3). Repeating the same arguments for a countable number of rational δ gives $\limsup_{k \rightarrow \infty} \phi_k \geq \phi_\alpha^*$. At the same time,

$\limsup_{k \rightarrow \infty} \phi_k \leq \lim_{k \rightarrow \infty} \Psi(\mathcal{S}_{k, a_{k+1}}^*) = \phi_\alpha^*$ μ -a.s. from Lemma 2, so that

$$\limsup_{k \rightarrow \infty} \phi_k = \phi_\alpha^*, \quad \mu\text{-a.s.} \quad (\text{B.4})$$

(iii) Assume that $\liminf_{k \rightarrow \infty} \phi_k < \phi_\alpha^* - \epsilon$ for some $\epsilon > 0$. We show that this event has probability zero.

First, we show that there exists K^* (μ -a.s.) such that for any $k > K^*$,

$$\phi_{k-1} < \phi_\alpha^* + \epsilon/6, \quad (\text{B.5})$$

$$(\epsilon/6) [\alpha/(1+a_k)] - A/(1+a_k)^2 > 0, \quad (\text{B.6})$$

$$\phi_k - \phi_{k-1} > -\epsilon/6. \quad (\text{B.7})$$

(B.5) follows from (B.4) and (B.6) from $a_k/k \rightarrow \alpha$ μ -a.s. **H_Φ4** implies that $F_\Phi(\xi_{k-1}, X_k) \geq -\delta_1(\underline{L}/2)$ for $k > K_0$ (since $\lambda_{\min}[\mathbf{M}(\xi_{k-1})] > \underline{L}/2$ for $k > K_0$) and (B.1) gives

$$\phi_k - \phi_{k-1} \geq -\frac{\delta_1(\underline{L}/2)}{1+a_k} - \frac{\delta_0(\underline{L}/4) + \delta_2(\underline{L}/4)\|f(X_k)\|^4}{2(1+a_k)^2}$$

for $k > K_1$. **H_μ2** implies $\text{Prob}\{\|f(X_k)\| > k^\gamma\} \leq \mu_4/k^{4\gamma}$ and therefore, from Borel-Cantelli, $\text{Prob}\{\|f(X_k)\| > k^\gamma \text{ infinitely often}\} = 0$ for any $\gamma > 1/4$. It implies that exists K_3 (μ -a.s.) such that for any $k > K_3$

$$\phi_k - \phi_{k-1} \geq -\frac{\delta_1(\underline{L}/2)}{1+a_k} - \frac{\delta_0(\underline{L}/4) + \delta_2(\underline{L}/4)k^{4\gamma}}{2(1+a_k)^2}.$$

Take $\gamma < 1/2$, (B.7) then follows from $a_k/k \rightarrow \alpha$ μ -a.s. We work conditionally on $k > K^*$ in the rest of the proof.

Take $\epsilon > 0$ arbitrarily small and define U_k^ϵ as the number of down-crossings by ϕ_k of the interval $(\phi_\alpha^* - \epsilon, \phi_\alpha^* - \epsilon/6)$ for $k > K^*$: $\liminf_{k \rightarrow \infty} \phi_k < \phi_\alpha^* - \epsilon$ thus implies $U_\infty^\epsilon = \infty$. Also define a previsible process (C_k) as follows: set C_k to one when ϕ_{k-1} gets above $\phi_\alpha^* - \epsilon/6$ (strictly) and turn it to zero when ϕ_{k-1}

gets below $\phi_\alpha^* - \epsilon$ (strictly). This defines a new process

$$V_k = \sum_{i=K^*}^k C_i(\phi_i - \phi_{i-1})$$

which satisfies

$$V_k < -(5\epsilon/6) U_k^\epsilon + \max\{0, \phi_k - (\phi_\alpha^* - \epsilon/6)\}. \quad (\text{B.8})$$

Define k_j, \dots, \bar{k}_j as the j th set of consecutive indices for which $C_i = 1$, that is, such that $C_{k_j-1} = 0, C_{k_j} = C_{k_j+1} = \dots = C_{\bar{k}_j} = 1, C_{\bar{k}_j+1} = 0$. Within this set, we define k_j^* as the largest index such that $\phi_{k-1} \leq \phi_\alpha^* - \epsilon/6$ for all $k_j^* \leq k \leq \bar{k}_j$. Note that necessarily $k_j^* > k_j$ (since $\phi_{k_j-1} > \phi_\alpha^* - \epsilon/6$) and $k_j^* < \bar{k}_j$ (since (B.7) is satisfied for $k > K^*$). We bound $\mathbb{E}\{V_{\bar{k}_j} - V_{k_j} | X_1^{k_j-1}\}$ as follows.

We have $\phi_{k_j^*-2} > \phi_\alpha^* - \epsilon/6$ (from the definition of k_j^*) and $\phi_{k_j^*-1} - \phi_{k_j^*-2} > -\epsilon/6$ from (B.7). Also, (B.5) implies $\phi_{k_j-1} < \phi_\alpha^* + \epsilon/6$. Therefore, $\phi_{k_j^*-1} - \phi_{k_j-1} > -\epsilon/2$. Now, for all k such that $k_j^* \leq k \leq \bar{k}_j$, $\phi_{k-1} \leq \phi_\alpha^* - \epsilon/6$ and thus $\mathbb{E}\{\phi_k - \phi_{k-1} | X_1^{k-1}\} > 0$ from (B.6) and (B.2). We obtain $\mathbb{E}\{V_{\bar{k}_j} - V_{k_j} | X_1^{k_j-1}\} > -\epsilon/2$ and $\mathbb{E}\{V_k | X_1^{K^*-1}\} > -(\epsilon/2)\mathbb{E}\{U_k^\epsilon | X_1^{K^*-1}\}$, which, together with (B.8) gives

$$\mathbb{E}\{U_k^\epsilon | X_1^{K^*-1}\} < \frac{3}{\epsilon} \mathbb{E}\{\max\{0, \phi_k - (\phi_\alpha^* - \epsilon/6)\} | X_1^{K^*-1}\} < \infty$$

where the last inequality follows from $\mathbb{E}\{|\phi_k|\} < \infty$. It implies $\text{Prob}\{U_\infty^\epsilon = \infty\} = 0$, that is, $\liminf_{k \rightarrow \infty} \phi_k > \phi_\alpha^* - \epsilon$, μ -a.s. Repeating the same arguments for a countable number of rational ϵ gives $\liminf_{k \rightarrow \infty} \phi_k = \phi_\alpha^*$, μ -a.s., which completes the proof. \square

Acknowledgements

This work originated from a comment by V.V. Fedorov (GlaxoSmithKline, USA) at a meeting on experimental design held in Cardiff in April 2000, see Pronzato (2001b); he is gratefully acknowledged for pointing out the connection between the results presented in this meeting and optimum design with constraints. Also, stimulating discussion with H.P. Wynn (Warwick University and London School of Economics, UK) and E. Thierry (Laboratoire I3S) on different occasions proved very helpful, and the comments of two anonymous referees have been extremely useful to improve the paper.

References

- Albright, S., Derman, C., 1972. Asymptotically optimal policies for the stochastic sequential assignment problem. *Management Science* 19 (1), 46–51.
- Box, G., Lucas, H., 1959. Design of experiments in nonlinear situations. *Biometrika* 46, 77–90.
- C.E.R.N., 1994. The compact muon solenoid, technical proposal, trigger and data acquisition. Tech. Rep. CERN/LHCC 94-38, European Laboratory for Particle Physics.
- Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. *Modelling Extremal Events*. Springer, Berlin.
- Fedorov, V., 1989. Optimal design with bounded density: optimization algorithms of the exchange type. *J. Statist. Planning and Inference* 22, 1–13.
- Fedorov, V., Hackl, P., 1997. *Model-Oriented Design of Experiments*. Springer, Berlin.

- Jennrich, R., 1969. Asymptotic properties of nonlinear least squares estimation. *Annals of Math. Stat.* 40, 633–643.
- Kiefer, J., Wolfowitz, J., 1960. The equivalence of two extremum problems. *Canadian Journal of Mathematics* 12, 363–366.
- Pronzato, L., 1999. Sequential selection of observations in randomly generated experiments. In: *Proc. ProbaStat'98, Smolenice (Slovaquie), Feb. 98, Tatra Mountains Mathematical Publications*. Vol. 17. pp. 167–175.
- Pronzato, L., 2001a. Optimal and asymptotically optimal decision rules for sequential screening and resource allocation. *IEEE Transactions on Automatic Control* 46 (5), 687–697.
- Pronzato, L., 2001b. Sequential construction of an experiment design from and i.i.d. sequence of experiments without replacement. In: Atkinson, A., Bogacka, B., Zhigljavsky, A. (Eds.), *Optimum Design 2000*. Kluwer, Dordrecht, Ch. 11, pp. 113–122.
- Pronzato, L., 2004. A minimax equivalence theorem for optimum bounded design measures. *Statistics & Probability Letters* 68, 325–331.
- Sahm, M., Schwabe, R., 2001. A note on optimal bounded designs. In: Atkinson, A., Bogacka, B., Zhigljavsky, A. (Eds.), *Optimum Design 2000*. Kluwer, Dordrecht, Ch. 13, pp. 131–140.
- Williams, D., 1991. *Probability with Martingales*. Cambridge University Press, Cambridge.
- Wu, C., Wynn, H., 1978. The convergence of general step-length algorithms for regular optimum design criteria. *Annals of Statistics* 6 (6), 1273–1285.
- Wynn, H., 1982. Optimum submeasures with applications to finite population sampling. In: Gupta, S., Berger, J. (Eds.), *Statistical Decision Theory and Related Topics III. Proc. 3rd Purdue Symp., vol. 2*. Academic Press, New York, pp. 485–495.